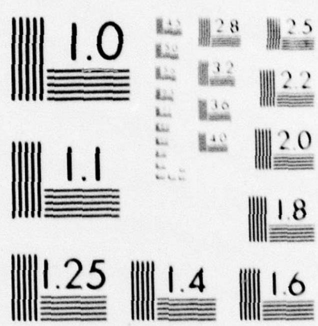
AD-A056 734 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
A NOTE ON OPTIMAL SUBSET SELECTION PROCEDURES. (U)
JUN 78 S S GUPTA, D HUANG N00014-75-C-0455
UNCLASSIFIED MMS-470-REV NL

1 OF 1
AD
A056 734

END
DATE
FILMED
9-78
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

LEVEL II

AD A056734

6

A NOTE ON OPTIMAL SUBSET SELECTION PROCEDURES

by

10 Shanti S. Gupta, Purdue University

and

Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

9 Mimeograph series

AD No.
DDC FILE COPY

DDC
JUL 26 1978
F

15 N00014-75-C-0455

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #470

Revised, Jun[redacted]78

11

12 10p.

14 MMS-470-REV

*This research was supported by Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

78 07 25 028

291730

JLB

A Note on Optimal Subset Selection Procedures*

by

Shanti S. Gupta, Purdue University

and

Deng-Yuan Huang, Academia Sinica, Taipei, Taiwan

1. Introduction. A common problem faced by an experimenter is one of comparing several categories or populations. The classical approach to this problem has been to use a test for homogeneity, i.e., to test whether all categories (populations) are identical or not. This approach is not always realistic and it is often inadequate. The inadequacy lies in the fact that only two decisions, accept or reject the hypothesis, are available to the experimenter. The experimenter is then faced with the problem of what to do next, especially if the hypothesis is rejected. These difficulties and inadequacies may be alleviated by formulating the problem as multiple decision problems aimed at selection or ranking (ordering) of the k populations. This has led to the rapid development of selection and ranking theory during the last two decades. Many reasonable rules have been proposed. Some desirable properties of these rules have been studied. However, very little work has been done to study the optimality of selection procedures, especially in the subset selection approach. In this paper, we are interested in deriving some methods to construct optimal subset selection procedures. Some classical selection rules are constructed as special cases.

Let $\pi_1, \pi_2, \ldots, \pi_k$ represent $k (\geq 2)$ populations (treatments) and let $X_{i1}, \ldots, X_{in_i}$ be n_i independent random observations from π_i. The quality of

*This research was supported by Office of Naval Research Contract N00014-75-C-0455 at Purdue University.

the ith population π_i is characterized by a real-valued parameter θ_i, usually unknown. Let $\Omega = \{\underline{\theta} | \underline{\theta}' = (\theta_1,\ldots,\theta_k)\}$ denote the whole parameter space. Let $\tau_{ij} = \tau_{ij}(\underline{\theta})$ be a measure of separation between π_i and π_j. We assume that there exists a monotonically nonincreasing function h such that $\tau_{ji} = h(\tau_{ij})$. Let $\Omega_i = \{\underline{\theta} | \tau_{ij} \geq \tau_0, j \neq i\}$, $1 \leq i \leq k$, and $\Omega_0 = \Omega - \bar{\Omega}$, where $\bar{\Omega} = \bigcup_{i=1}^{k} \Omega_i$. Thus $\Omega = \bigcup_{i=0}^{k} \Omega_i$ is a partition of Ω. For this problem, we assume τ_0 and τ_{ii} as known with $\tau_0 > \tau_{ii}$ for all i. Let $\tau_i = \min_{j \neq i} \tau_{ij}$, $1 \leq i \leq k$. We define $\tau^* = \max_{1 \leq \ell \leq k} \tau_\ell$. The population associated with τ^* will be called the best population. It should be pointed out that if $\underline{\theta} \in \Omega_i$, then $\tau_i \geq \tau_j$ for all j, since for some j, $j \neq i$, $\tau_{ji} = h(\tau_{ij}) \leq h(\tau_0) \leq h(\tau_{ii}) = \tau_{ii} < \tau_0$. Thus if $\underline{\theta} \in \Omega_i$, then π_i is the best population. A selection of a subset containing the best population is called a correct selection (CS). (Note that in case of ties any one of the best populations corresponding to τ^* is "tagged" as the best population.) To illustrate the above notation, we assume that the observations are drawn from π_i which has a normal distribution with unknown mean θ_i $(i = 1,\ldots,k)$ and known variance σ^2. We can define $\tau_{ij} = \theta_i - \theta_j$; then it can be seen that $\tau_i = \theta_i - \theta_{[k]}$ if $\theta_i < \theta_{[k]}$ and $\tau_i = \theta_i - \theta_{[k-1]}$ if $\theta_i = \theta_{[k]}$, where $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$. In this case, $\tau_{ii} = 0$ for all i. Thus the population with the largest mean, $\theta_{[k]}$, is the best. If instead $\tau_{ij} = \theta_j - \theta_i$ then the population with the smallest mean $\theta_{[1]}$ would be the best. In the above example, we have $h(t) = -t$ which is a decreasing function.

Let the observed sample vector be denoted by $\underline{X}' = (\underline{X}_1',\ldots,\underline{X}_k')$ where $\underline{X}_i$ has components $X_{i1},\ldots,X_{in_i}$, $i = 1,\ldots,k$. Let $\delta = (\delta_1,\ldots,\delta_k)$ be a selection procedure where $\delta_i(\underline{x})$ is the probability of selecting π_i $(1 \leq i \leq k)$ based on the observed vector $\underline{X} = \underline{x}$ from the k populations. As measures of goodness or optimality of a selection rule, consider two quantities (cf. Lehmann [1]) $R(\underline{\theta},\delta)$ and $S(\underline{\theta},\delta)$. We define $S(\underline{\theta},\delta) = P_{\underline{\theta}}(CS|\delta)$ and $R(\underline{\theta},\delta) = \sum_{i=1}^{k} R^{(i)}(\underline{\theta},\delta_i)$, where $R^{(i)}(\underline{\theta},\delta_i) = P\{\text{Selecting } \pi_i|\delta\}$. Let S be the size of the selected subset. Thus $R(\underline{\theta},\delta) = E(S|\delta)$. For a specified γ, $(0 < \gamma < 1)$, we may restrict our attention to the class $\mathscr{C}$ of all δ such that

$$(1) \qquad S(\underline{\theta},\delta) \geq \gamma \text{ for } \underline{\theta} \in \bar{\Omega}.$$

We are interested in constructing an optimal procedure δ^0 in $\mathscr{C}$ which minimizes the supremum of $R(\underline{\theta},\delta)$ over Ω for all $\delta \in \mathscr{C}$, i.e.,

$$(2) \qquad \sup_{\underline{\theta}\in\Omega} R(\underline{\theta},\delta^0) = \min_{\delta\in\mathscr{C}} \sup_{\underline{\theta}\in\Omega} R(\underline{\theta},\delta).$$

We restrict attention to those selection procedures which depend upon the observations only through a sufficient and *maximal invariant statistic* Z_{ij} which are defined as follows:

$$Z_{ij} = f(X_{i1},\ldots,X_{in_i}; X_{j1},\ldots,X_{jn_j}).$$

This Z_{ij} is based on the n_i and n_j observations from π_i and π_j $(i,j = 1,2,\ldots,k)$, respectively. It is well known that the distribution of Z_{ij} depends only on τ_{ij}. For any i, let the joint density of Z_{ij}, $j \neq i$, be $p_{\underline{\theta}}(\underline{z}_i)$. Let $p_{\underline{\theta}}(\underline{z}_i)$ be denoted by $p_0(\underline{z}_i)$ when $\tau_{i1} = \ldots = \tau_{ik} = \tau_{ii}$ = constant and by $p_i(\underline{z}_i)$ when $\tau_{i1} = \ldots = \tau_{ik} = \tau_0$, $1 \leq i \leq k$. In a given problem the function f is so chosen as to indicate the measure of separation between the populations in a reasonable way. In case of the above normal means example, a choice of Z_{ij} might be $\bar{X}_i - \bar{X}_j$, where $\bar{X}_i = \frac{1}{n_i}\sum_{\ell=1}^{n_i} X_{i\ell}$ and $\bar{X}_j = \frac{1}{n_j}\sum_{\ell=1}^{n_j} X_{j\ell}$. Let ν be a σ-finite measure on R^{k-1}.

Oosterhoff [2] defines a monotone likelihood ratio for a random vector $\underline{X}$ with m components as follows: Let $\underline{\theta}$ be an m-dimensional vector of parameters. A partial ordering of points in R^m is defined by $\underline{x}_1 \prec \underline{x}_2$, $\underline{x}_i' = (x_{i1},\ldots,x_{im})$, $i = 1,2$, meaning that $x_{1j} \le x_{2j}$ for $j = 1,2,\ldots,k$, and the inequality is strict for at least one component. *The density* $f_{\underline{\theta}}(\underline{x})$ *has monotone likelihood ratio* (MLR) if for all $\underline{\theta}_1 \prec \underline{\theta}_2$, $f_{\underline{\theta}_2}(\underline{x})/f_{\underline{\theta}_1}(\underline{x})$ is nondecreasing in $\underline{x}$.

Now we state and prove a theorem which provides a solution to the restricted minimax problem as stated in (1) and (2) (cf Lehmann [1]).

Theorem: Suppose that the density $p_{\underline{\theta}}(\underline{z})$ has the MLR property. If $R(\underline{\theta},\delta^0)$ is maximized at $\tau_{ij} = \tau_{ii}$ = constant, for all i,j, where δ^0 is given by

$$\delta_i^0(\underline{z}_i) = \begin{cases} 1 & \text{if } p_i(\underline{z}_i) > c\, p_0(\underline{z}_i), \\ \lambda_i & \qquad = \quad , \\ 0 & \qquad < \quad , \end{cases}$$

such that $c(>0)$ and λ_i are determined by $\int \delta_i^0 p_i = \gamma$, $1 \le i \le k$. Then $\delta^0 = (\delta_1^0,\ldots,\delta_k^0)$ minimizes $\sup_{\underline{\theta}\in\Omega} R(\underline{\theta},\delta)$ subject to $\inf_{\underline{\theta}\in\Omega} S(\underline{\theta},\delta) \ge \gamma$.

Proof. For any $\delta \in \mathscr{C}$,

$\underline{\theta} \in \bar{\Omega}$ implies $\underline{\theta} \in \Omega_i$ for some i, thus

$$S(\underline{\theta},\delta) = \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i) d\nu(\underline{z}_i) \ge \min_{1\le i\le k} \inf_{\underline{\theta}\in\Omega_i} \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i) d\nu(\underline{z}_i).$$

We have

$$\inf_{\underline{\theta}\in\bar{\Omega}} S(\underline{\theta},\delta) = \min_{1\le i\le k} \inf_{\underline{\theta}\in\Omega_i} \int \delta_i(\underline{z}_i) p_{\underline{\theta}}(\underline{z}_i) d\nu(\underline{z}_i).$$

Hence for any $\delta \in \mathscr{C}$, $\inf_{\underline{\theta}\in\Omega_i} \int \delta_i(\underline{z}_i)\, p_{\underline{\theta}}(\underline{z}_i) d\nu(\underline{z}_i) \ge \gamma$, $1 \le i \le k$, and by the assumption that $\int \delta_i^0 p_i = \gamma$, it follows that

$$\int(\delta_i-\delta_i^0)(p_i-cp_0) \leq 0$$

which implies

$$\int\delta_i^0\, p_0 \leq \int\delta_i p_0.$$

By the assumption, $\delta_i^0(\underline{z}_i)$ is nondecreasing in $\underline{z}_i$, we have

$$\inf_{\underline{\theta}\in\Omega} S(\underline{\theta},\delta^0) = \min_{1\leq i\leq k} \int\delta_i^0 p_i = \gamma.$$

If $R(\underline{\theta},\delta^0)$ is maximized at $\tau_{ij} = \tau_{ii}$ = constant, for all i,j, then

$$\sup_{\underline{\theta}\in\Omega} R(\underline{\delta},\delta) \geq \sum_{i=1}^{k} \int\delta_i p_0 \geq \sum_{i=1}^{k} \int\delta_i^0 p_0 = \sup_{\underline{\theta}\in\Omega} R(\underline{\theta},\delta^0),$$

which completes the proof.

Example: Let $\pi_1,\pi_2,\ldots,\pi_k$ be k independent normal populations with means $\theta_1,\ldots,\theta_k$ and common variance $\sigma^2 = 1$. Our interest is to select a nonempty subset of the k populations containing the best. The ordered θ_i's are denoted by $\theta_{[1]} \leq\ldots\leq \theta_{[k]}$. It is assumed that there is no prior knowledge of the correct pairing of the ordered and the unordered θ_i's. Our goal is to select a nonempty subset of the k populations so as to include the population associated with $\theta_{[k]}$.

Let $\bar{X}_i$, $1 \leq i \leq k$, denote the sample means of independent samples of size n from these populations. Let the joint likelihood function of $\bar{X}_i$, $i=1,2,\ldots,k$, be

$$g_{\underline{\theta}}(\underline{x}) = \prod_{j=1}^{k} g_{\theta_i}(\bar{x}_i),$$

where
$$g_{\theta_i}(\bar{x}_i) = \frac{\sqrt{n}}{\sqrt{2\pi}}\, e^{-\frac{n}{2}(\bar{x}_i-\theta_i)^2}, \quad 1 \leq i \leq k.$$ Let

$\tau_{ij} = \tau_{ij}(\underline{\theta}) = \theta_i-\theta_j$, $1 \leq j \leq k$, $j \neq i$, $\tau_{ii} = 0$, $\tau_0 = \Delta > 0$ and $Z_{ij} = \bar{X}_i-\bar{X}_j$, $j \neq i$. Let $\underline{z}_i' = (z_{i1},\ldots,z_{ik})$ and $\underline{\tau}_i' = (\tau_{i1},\ldots,\tau_{ik})$.

$$p_{\underline{\theta}}(\underline{z}_i) = (2\pi)^{\frac{k-1}{2}} |\Sigma|^{-1/2} \exp\{-(\underline{z}_i-\underline{\tau}_i)'\Sigma^{-1}(\underline{z}_i-\underline{\tau}_i)\},$$

where $\Sigma_{(k-1)x(k-1)} = \frac{1}{n}\begin{pmatrix} 2 & & 1 \\ & \cdot & \\ 1 & & .2 \end{pmatrix}$ is the positive definite covariance matrix of Z_{ij}'s. We know that

$$\frac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} = \exp\{\underline{z}_i'\Sigma^{-1}\underline{\Delta} + \underline{\Delta}'\Sigma^{-1}\underline{z}_i - \underline{\Delta}'\Sigma^{-1}\underline{\Delta}\}$$

is nondecreasing in z_{ij}, $j \neq i$, where $\underline{\Delta}' = ($ $,\Delta)$. And

$$\frac{p_i(\underline{z}_i)}{p_0(\underline{z}_i)} > c$$

is equivalent to

(3) $$\bar{x}_i > \frac{1}{k-1} \sum_{j\neq i} \bar{x}_j + d.$$

For any i, let $\mu_i = \sum_{j\neq i} (\theta_j - \theta_i)$, we have

$$\mu_i = \mu - k\theta_i$$

where $\mu = \sum_{j=1}^{k} \theta_j$. Hence, $\mu_1 = \mu_2 = \ldots = \mu_k$ if and only if $\theta_1 = \theta_2 = \ldots = \theta_k$. We order μ_i's as $\mu_{[1]} \leq \ldots \leq \mu_{[k]}$. Since

$$\begin{aligned} R(\underline{\theta},\delta) &= \sum_{i=1}^{k} P\{\bar{X}_i > \frac{1}{k-1} \sum_{j\neq i} \bar{X}_j + d\} \\ &= \sum_{i=1}^{k} \int \Phi\{\sqrt{k-1}(y - \frac{1}{k-1}\mu_i - d)\} d\Phi(y) \\ &= f(\mu_1,\ldots,\mu_k) \\ &\leq f(\mu_{[1]},\ldots,\mu_{[1]}) \\ &= f(\theta_{[k]},\ldots,\theta_{[k]}) \end{aligned}$$

$$= f(0,\ldots,0)$$

$$= R(\underline{0},\delta),$$

hence, it follows that the supremum of $R(\underline{\theta},\delta^0)$ over Ω occurs at $\theta_1 = \ldots = \theta_k$, i.e., $\tau_{ij} = \tau_{ii} = 0$. Thus the result of the theorem can be applied.

Note that the above procedure (3) is a rule of the type proposed by Seal [3] to select a subset containing the population associated with the largest θ_i's.

Acknowledgment

The authors wish to thank the referees for their valuable comments and suggestions which have improved and simplified the presentation.

References

[1] Lehmann, E. L. (1961). Some model I problems of selection. Ann. Math. Statist. 32, 990-1012.

[2] Oosterhoff, J. (1969). Combination of One-Sided Statistical Tests. Mathematical Centre Tracts, Amsterdam.

[3] Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. 36, 387-397.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #470 (Revised)	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Note on Optimal Subset Selection Procedures		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #470
7. AUTHOR(s) Shanti S. Gupta and D. Y. Huang		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Purdue University W. Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE June 1978
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

optimal selection, restricted minimax, monotone likelihood ratio,

concerns

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, we are concerned with the construction of "optimal" subset selection procedures. Some classical selection procedures are considered as special cases.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)